

EKIPO 5

ALGORITMOS GENETICOS

Los algoritmos genéticos consisten en generar una serie de posibles soluciones al azar y en ir creando sucesivas generaciones (soluciones hijas, mezcla de otras soluciones posibles anteriores), dando más importancia y más peso a las mejores soluciones.

El objetivo de los AG es buscar dentro de un espacio de hipótesis candidatas la mejor de ellas. En los AG la mejor hipótesis es aquella que optimiza a una métrica predefinida para el problema dado, es decir, la que más se aproxima a dicho valor numérico una vez evaluada por la función de evaluación.

Algoritmos Genéticos Generacionales

Se asemejan a la forma de reproducción de los insectos, donde una generación pone huevos, se aleja geográficamente o muere y es substituida por una nueva. En este momento se realizan cruces en una piscina de individuos, los descendientes son puestos en otra, al final de la fase reproductiva se elimina la generación anterior y se pasa a utilizar la nueva. Este modelo también es conocido como Algoritmo Genético Canónico.

1.1. Antecedentes

El algoritmo genético es una técnica de búsqueda basada en la teoría de la evolución de Darwin, que ha cobrado tremenda popularidad en todo el mundo durante los últimos años. Se presentarán aquí los conceptos básicos que se requieren para abordarla, así como unos sencillos ejemplos que permitan a los lectores comprender cómo aplicarla al problema de su elección.

En los últimos años, la comunidad científica internacional ha mostrado un creciente interés en una nueva técnica de búsqueda basada en la teoría de la evolución y que se conoce como el **algoritmo genético**. Esta técnica se basa en los mecanismos de selección que utiliza la naturaleza, de acuerdo a los cuales los individuos más aptos de una población son los que sobreviven, al adaptarse más fácilmente a los cambios que se producen en su entorno. Hoy en día se sabe que estos cambios se efectúan en los genes de un individuo (unidad básica de codificación de cada uno de los atributos de un ser vivo), y que sus atributos más deseables (i.e., los que le permiten adaptarse mejor a su entorno) se transmiten a sus descendientes cuando éste se reproduce sexualmente.

1.2. Definición

Los Algoritmos Genéticos (AGs) son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización. Están basados en el proceso genético de los organismos vivos. A lo largo de las generaciones, las poblaciones evolucionan en la naturaleza de acorde con los principios de la selección natural y la supervivencia de los más fuertes, postulados por Darwin. Por imitación de este proceso, los Algoritmos Genéticos son capaces de ir creando soluciones para problemas del mundo real. La evolución de dichas soluciones hacia valores óptimos del problema depende en buena medida de una adecuada codificación de las mismas.

Un algoritmo genético consiste en una función matemática o una rutina de software que toma como entradas a los ejemplares y retorna como salidas cuales de ellos deben generar descendencia para la nueva generación.

1.5. Limitaciones

El poder de los Algoritmos Genéticos proviene del hecho de que se trata de una técnica robusta, y pueden tratar con éxito una gran variedad de problemas provenientes de diferentes áreas, incluyendo aquellos en los que otros métodos encuentran dificultades. Si bien no se garantiza que el Algoritmo Genético encuentre la solución óptima, del problema, existe evidencia empírica de que se encuentran soluciones de un nivel aceptable, en un tiempo competitivo con el resto de algoritmos de optimización combinatoria. En el caso de que existan técnicas especializadas para resolver un determinado problema, lo más probable es que superen al Algoritmo Genético, tanto en rapidez como en eficacia. El gran campo de aplicación de los Algoritmos Genéticos se relaciona con aquellos problemas para los cuales no existen técnicas especializadas. Incluso en el caso en que dichas técnicas existan, y funcionen bien, pueden efectuarse mejoras de las mismas hibridándolas con los Algoritmos Genéticos.

1.6. Como Saber si es Posible usar un Algoritmo Genético

La aplicación más común de los algoritmos genéticos ha sido la solución de problemas de optimización, en donde han mostrado ser muy eficientes y confiables. Sin embargo, no todos los problemas pudieran ser apropiados para la técnica, y se recomienda en general tomar en cuenta las siguientes características del mismo antes de intentar usarla:

- Su espacio de búsqueda (i.e., sus posibles soluciones) debe estar delimitado dentro de un cierto rango.
- Debe poderse definir una función de aptitud que nos indique qué tan buena o mala es una cierta respuesta.
- Las soluciones deben codificarse de una forma que resulte relativamente fácil de implementar en la computadora

Los algoritmos genéticos se revalorizaron ya que poseen las siguientes ventajas competitivas:

- **Solo necesitan asesoramiento del experto cuando se agregan o suprimen variables al modelo. Los Sistemas Expertos requieren la presencia del mismo ante cada modificación del entorno.**
- **Los algoritmos genéticos solo requieren el asesoramiento del experto para identificar las variables pertinentes , aunque no es necesario que éstos definan sus valores ni sus relaciones (las reglas) iniciales o finales. Los Sistemas Expertos solo trabajan con las reglas y valores que les dictan los seres humanos**

3.7. Algoritmos Genéticos Paralelos

En este apartado se introducirán tres maneras diferentes de explotar el paralelismo de los Algoritmos Genéticos, por medio de los denominados modelos de islas. Para una profundización sobre el tema puede consultarse Stender.

Modelos de islas.

La idea básica consiste en dividir la población total en varias subpoblaciones en cada una de las cuales se lleva, a cabo un Algoritmo Genético. Cada cierto número de generaciones, se efectúa un intercambio de información entre las subpoblaciones, proceso que se denomina migración. La introducción de la migración hace que los modelos de islas sean capaces de explotar las diferencias entre las diversas subpoblaciones, obteniéndose de esta manera una fuente de diversidad genética. Cada subpoblación es una "isla", definiéndose un procedimiento por medio del cual se mueve el material genético de una "isla" a otra. La determinación de la tasa de migración, es un asunto de capital importancia, ya que de ella puede depender la convergencia prematura de la búsqueda.

Se pueden distinguir diferentes modelos de islas en función de la comunicación entre las subpoblaciones. Algunas comunicaciones típicas son las siguientes:

- Comunicación en estrella, en la cual existe una subpoblación que es seleccionada como maestra (aquella que tiene mejor media en el valor de la función objetivo), siendo las demás consideradas como esclavas. Todas las subpoblaciones esclavas mandan sus h_1 mejores individuos ($h_1 > 1$) a la subpoblación maestra la cual a su vez manda sus h_2 mejores individuos ($h_2 > 1$) a cada una de las subpoblaciones esclavas. Véase Figura 12.

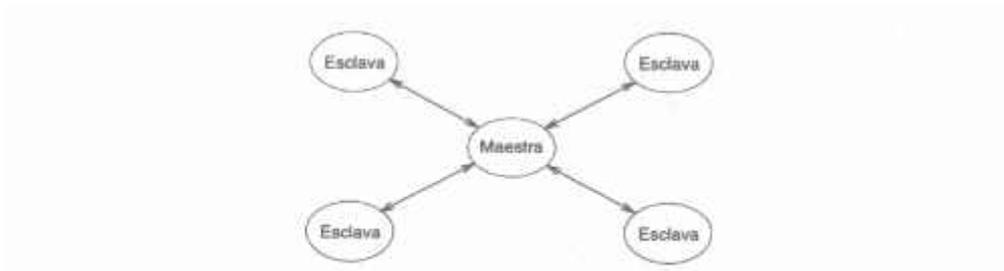


Figura 12: Algoritmo Genético Paralelo. Modelo de Islas. Comunicación en estrella

Figura 12

- Comunicación en red, en la cual no existe una jerarquía entre las subpoblaciones, mandando todas y cada una de ellas sus h_3 ($h_3 > 1$) mejores individuos al resto de las subpoblaciones. Véase Figura 13.

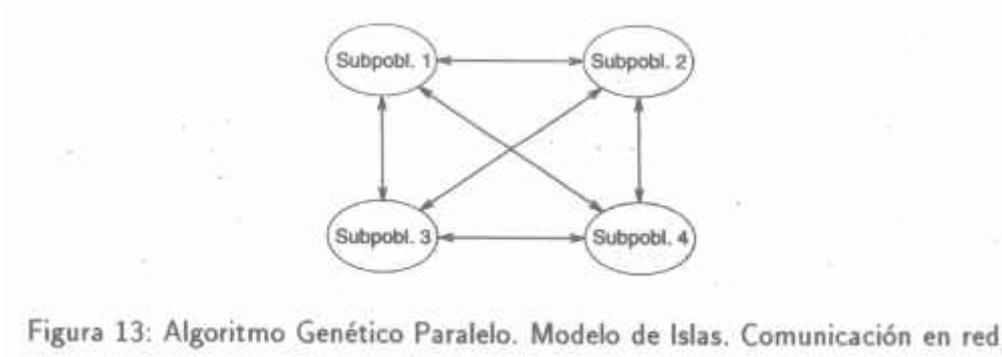
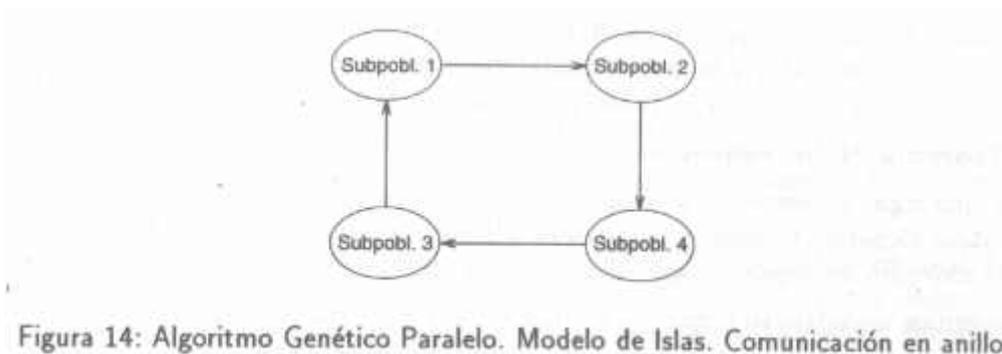


Figura 13

- Comunicación en anillo, en la cual cada subpoblación envía sus h_4 mejores individuos ($h_4 > 1$), a una población vecina, efectuándose la migración en un único sentido de flujo. Véase Figura 14.

El modelo de islas ha sido utilizado por varios autores (Whitley y Starkweather, Gorges-Schleuter, Tanese).



CLASES DE ALGORITMOS GENÉTICOS

Algoritmos Genéticos Generacionales

Se asemejan a la forma de reproducción de los insectos, donde una generación pone huevos, se aleja geográficamente o muere y es substituida por una nueva. En este momento se realizan cruces en una piscina de individuos, los descendientes son puestos en otra, al final de la fase reproductiva se elimina la generación anterior y se pasa a utilizar la nueva. Este modelo también es conocido como Algoritmo Genético Canónico.

Algoritmos Genéticos de estado Fijo.

Utilizan el esquema generacional de los mamíferos y otros animales de vida larga,

donde coexisten padres y sus descendientes, permitiendo que los hijos sean educados por sus progenitores, pero también que a la larga se genere competencia entre ellos. En este modelo, no solo se deben seleccionar los dos individuos a ser padres, si no también cuales de la población anterior serán eliminados, para dar espacio a los descendientes.

La diferencia esencial entre el reemplazo generacional y el modelo de estado fijo es que las estadísticas de la población son recalculadas luego de cada cruce y los nuevos descendientes están disponibles inmediatamente para la reproducción. Esto permite al modelo utilizar las características de un individuo prometedor tan pronto como es creado.

Algunos autores dicen que este modelo tiende a evolucionar mucho más rápido que el modelo generacional, sin embargo investigaciones de Goldberg y Deb [GOLDBERG 93], encontraron que las ventajas parecen estar relacionadas con la alta tasa de crecimiento inicial, ellos dicen que los mismos efectos pueden ser obtenidos en rangos de adaptación exponencial o selección por competencia. No encontraron evidencia que este modelo sea mejor que el Generacional.

Algoritmos Genéticos Paralelos.

Parte de la metáfora biológica que motivo a utilizar la búsqueda genética consiste en que es inherentemente paralela, donde al evolucionar se recorren simultáneamente muchas soluciones, cada una representada por un individuo de la población. Sin embargo, es muy común en la naturaleza que no solo sea una población evolucionando, si no varias poblaciones, normalmente aisladas geográficamente, que originan respuestas diferentes a la presión evolutiva. Esto origina dos modelos que toman en cuenta esta variación, y utilizan no una población como los anteriores si no múltiples concurrentemente.

Modelos de Islas.

Si se tiene una población de individuos, esta se divide en subpoblaciones que evolucionan independientemente como un Algoritmo Genético normal.

Ocasionalmente, se producen migraciones entre ellas, permitiéndoles intercambiar material genético.

Con la utilización de la migración, este modelo puede explotar las diferencias en las subpoblaciones; esta variación representa una fuente de diversidad genética. Sin embargo, si un número de individuos emigran en cada generación, ocurre una mezcla global y se eliminan las diferencias locales, y si la migración es infrecuente, es probable que se produzca convergencia prematura en las subpoblaciones.

Modelo Celular

Coloca cada individuo en una matriz, donde cada uno sólo podrá buscar reproducirse con los individuos que tenga a su alrededor (mas cerca de casa) escogiendo al azar o al mejor adaptado. El descendiente pasara a ocupar una posición cercana.

No hay islas en este modelo, pero hay efectos potenciales similares. Asumiendo que el cruce esta restringido a individuos adyacentes, dos individuos separados por 20 espacios están tan aislados como si estuvieran en dos islas, este tipo de separación es conocido como aislamiento por distancia.

Luego de la primera evaluación, los individuos están todavía distribuidos al azar sobre la matriz. Posteriormente, empiezan a emerger zonas como cromosomas y adaptaciones semejantes. La reproducción y selección local crea tendencias evolutivas aisladas, luego

de varias generaciones, la competencia local resultara en grupos mas grandes de individuos semejantes.

Evaluación y selección

Durante la evaluación, se decodifica el gen, convirtiéndose en una serie de parámetros de un problema, se halla la solución del problema a partir de esos parámetros, y se le da una puntuación a esa solución en función de lo cerca que esté de la mejor solución. A esta puntuación se le llama *fitness*.

Por ejemplo, supongamos que queremos hallar el máximo de la función $f(x) = -x^2 + 2x$, una parábola invertida con el máximo en $x=1$. En este caso, el único parámetro del problema es la variable x . La optimización consiste en hallar un x tal que $F(x)$ sea máximo. Crearemos, pues, una población de cromosomas, cada uno de los cuales contiene una codificación binaria del parámetro x . Lo haremos de la forma siguiente: cada byte, cuyo valor está comprendido entre 0 y 255, se transformará para ajustarse al intervalo $[-1,1]$, donde queremos hallar el máximo de la función.

Valor binario	Decodificación	Evaluación $f(x)$
10010100	21	0,9559
10010001	19	0,9639
101001	-86	0,2604
1000101	-58	0,6636

El fitness determina siempre los cromosomas que se van a reproducir, y aquellos que se van a eliminar, pero hay varias formas de considerarlo para seleccionar la población de la siguiente generación:

- Usar el orden, o rango, y hacer depender la probabilidad de permanencia o evaluación de la posición en el orden.
- Aplicar una operación al fitness para escalarlo; como por ejemplo el *escalado*

sigma. En este esquema el fitness se escala

- En algunos casos, el fitness no es una sola cantidad, sino diversos números, que tienen diferente consideración. Basta con que tal fitness forme un orden parcial, es decir, que se puedan comparar dos individuos y decir cuál de ellos es mejor. Esto suele suceder cuando se necesitan optimizar varios objetivos.

Una vez evaluado el fitness, se tiene que crear la nueva población teniendo en cuenta que los *buenos* rasgos de los mejores se transmitan a esta. Para ello, hay que seleccionar a una serie de individuos encargados de tan ardua tarea. Y esta selección, y la consiguiente reproducción, se puede hacer de dos formas principales:

- *Basado en el rango*: en este esquema se mantiene un porcentaje de la población, generalmente la mayoría, para la siguiente generación. Se coloca toda la población por orden de fitness, y los M menos dignos son eliminados y sustituidos por la descendencia de alguno de los M mejores con algún otro individuo de la población. A este esquema se le pueden aplicar otros criterios; por ejemplo, se crea la descendencia de uno de los paladines/amazonas, y esta sustituye al más parecido entre los perdedores. Esto se denomina *crowding*, y fue introducido por **DeJong**. En nuestro caso, se eliminaría el cromosoma número 3, y se sustituiría por un descendiente del cromosoma número 2 y otro aleatorio, escogido entre el 1 y el 4. En realidad, para este esquema se escoge un *crowding factor*, CF . Cuando nace una nueva criatura, se seleccionan CF individuos de la población, y se elimina al más parecido a la nueva criatura. Una variante de este es el muestreo estocástico universal, que trata de evitar que los individuos con más fitness copen la población; en vez de dar la vuelta a una ruleta con una ranura, da la vuelta a la ruleta con N ranuras, tantas como la población; de esta forma, la distribución estadística de descendientes en la nueva población es más parecida a la real.
- *Rueda de ruleta*: se crea un *pool* genético formado por cromosomas de la generación actual, en una cantidad proporcional a su fitness. Si la proporción hace que un individuo domine la población, se le aplica alguna operación de escalado. Dentro de este *pool*, se cogen parejas aleatorias de cromosomas y se emparejan, sin importar incluso que sean del mismo progenitor (para eso están otros operadores, como la mutación). Hay otras variantes: por ejemplo, en la nueva generación se puede incluir el mejor representante de la generación actual. En este caso, se denomina método *elitista*.
- *Selección de torneo*: se escogen aleatoriamente un número T de individuos de la población, y el que tiene puntuación mayor se reproduce, sustituyendo su descendencia al que tiene menor puntuación.

4.6 Crossover

Consiste en el intercambio de material genético entre dos cromosomas (a veces más, como el *operador orgía* propuesto por Eiben et al.). El *crossover* es el principal operador genético, hasta el punto que se puede decir que no es un algoritmo genético si no tiene *crossover*, y, sin embargo, puede serlo perfectamente sin mutación, según descubrió Holland. El *teorema de los esquemas* confía en él para hallar la mejor solución a un problema, combinando soluciones parciales.

Para aplicar el crossover, entrecruzamiento o recombinación, se escogen aleatoriamente dos miembros de la población. No pasa nada si se emparejan dos descendientes de los mismos padres; ello garantiza la perpetuación de un individuo con buena puntuación (y, además, algo parecido ocurre en la realidad; es una práctica utilizada, por ejemplo, en la cría de ganado, llamada *inbreeding*, y destinada a potenciar ciertas características frente a otras). Sin embargo, si esto sucede demasiado a menudo, puede crear problemas: toda la población puede aparecer dominada por los descendientes de algún gen, que, además, puede tener caracteres no deseados. Esto se suele denominar en otros métodos de optimización *atranque en un mínimo local*, y es uno de los principales problemas con los que se enfrentan los que aplican algoritmos genéticos.

En cuanto al teorema de los esquemas, se basa en la noción de *bloques de construcción*. Una buena solución a un problema está constituida por unos buenos bloques, igual que una buena máquina está hecha por buenas piezas. El crossover es el encargado de mezclar bloques buenos que se encuentren en los diversos progenitores, y que serán los que den a los mismos una buena puntuación. La presión selectiva se encarga de que sólo los buenos bloques se perpetúen, y poco a poco vayan formando una buena solución. El *teorema de los esquemas* viene a decir que la cantidad de *buenos bloques* se va incrementando con el tiempo de ejecución de un algoritmo genético, y es el resultado teórico más importante en algoritmos genéticos.

El intercambio genético se puede llevar a cabo de muchas formas, pero hay dos grupos principales

- *Crossover n-puntos*: los dos cromosomas se cortan por n puntos, y el material genético situado entre ellos se intercambia. Lo más habitual es un crossover de un punto o de dos puntos.

- *Crossover uniforme*: se genera un patrón aleatorio de 1s y 0s, y se intercambian los bits de los dos cromosomas que coincidan donde hay un 1 en el patrón. O bien se genera un número aleatorio para cada bit, y si supera una determinada probabilidad se intercambia ese bit entre los dos cromosomas.

- *Crossover especializados*: en algunos problemas, aplicar aleatoriamente el crossover da lugar a cromosomas que codifican soluciones inválidas; en este caso hay que aplicar el crossover de forma que genere siempre soluciones

válidas. Un ejemplo de estos son los operadores de crossover usados en el problema del viajante.

4.7 Mutación

En la Evolución, una mutación es un suceso bastante poco común (sucede aproximadamente una de cada mil replicaciones), como ya se ha visto anteriormente. En la mayoría de los casos las mutaciones son letales, pero en promedio, contribuyen a la diversidad genética de la especie. En un algoritmo genético tendrán el mismo papel, y la misma frecuencia (es decir, muy baja).

Una vez establecida la frecuencia de mutación, por ejemplo, uno por mil, se examina cada bit de cada cadena cuando se vaya a crear la nueva criatura a partir de sus padres (normalmente se hace de forma simultánea al crossover). Si un número generado aleatoriamente está por debajo de esa probabilidad, se cambiará el bit (es decir, de 0 a 1 o de 1 a 0). Si no, se dejará como está. Dependiendo del número de individuos que haya y del número de bits por individuo, puede resultar que las mutaciones sean extremadamente raras en una sola generación.

No hace falta decir que no conviene abusar de la mutación. Es cierto que es un mecanismo generador de diversidad, y, por tanto, la solución cuando un algoritmo genético está estancado, pero también es cierto que reduce el algoritmo genético a una búsqueda aleatoria. Siempre es más conveniente usar otros mecanismos de generación de diversidad, como aumentar el tamaño de la población, o garantizar la aleatoriedad de la población inicial.

Este operador, junto con la anterior y el método de selección de ruleta, constituyen un *algoritmo genético simple*, sga, introducido por Goldberg en su libro.